

Sequencing of 5' end of phaSR<sup>III</sup>  
w/ Sequinase Kit

Purpose: Determine if phaSR<sup>III</sup> encodes  
a known protein (e.g. CD36 or  
Limp<sup>II</sup>)

Followed Kit instructions. Used primers:

T7 - on plasmid pcDNA1

oSA1.3 - in cDNA sequence

5' CTG TCG CTG TCC CCC TTC AG 3'

T7

short read

GGTACCGAGCTCGATCCACTAGTAACGGCCGCCAGTGTGCTCTAAGGCCACCTGCAGGGCTACTG  
CTGCTCCGGCCACTGCCTGAGACTCACCTTGCTGGAACGTGAGCCTCGGCTTCTGTCATCTCTG

long read

ACTCACCTTGCTGGAACGTGAGCCTCGGCTTCTGTCATCTCTGTGGCCTCTGTCGCTTCTGTCGCT  
GTCCCCCTTNAAGTCCCTGAGCCCCGCGAGCCCCGGGCCGCACACGCGACATGGGCGGCANNCCAGGG

mix

GGTACCGAGCTCGATCCACTAGTAACGGCCGCCAGTGTGCTCTAAGGCCACCTGCAGGGCTACTG  
CTGCTCCGGCCACTGCCTGAGACTCACCTTGCTGGAACGTGAGCCTCGGCTTCTGTCATCTCTGTG  
GCCTCTGTCGCTTCTGTCGCTGTCCCCCTTNAAGTCCCTGAGCCCCGCGAGCCCCGGGCCGCACACGC  
GACATGGGCGGCANNCCAGGG

OSA3.1

short read

CCGCACACGCGACATGGGCGGCACGGCCAGGGCGCTGGGTGGCGGTGGGGCTGGGCGTCGTGGGGC  
TGCTGTGCGCTGTGCTCGGTGTGGTTATGATCCTCGTGATGCCCTCGCTCATCAAACAGCAGGTAC  
TGAAGAATGTC (C?) GCATAGACCCAGCAGCCTGTCTTTGCAATGTGGAAGGAGA

long read

TGTGCTCGGTGTGGTTATGATCCTCGTGATGCCCTCGCTCATCAAACAGCAGGTACTGAAGAATGT  
CCGCATAGACCCAGCAGCCTGTCCTTTGCAATGTGGAAGGAGATCCCTGTACCCTTCTACTTGTG  
CGTCTACTTCTTCGAGGTGGTCAATCCCAGCGAGATCCTAAAGGGTGAGAA

mix T7 and OSA3.1

GGTACCGAGCTCGATCCACTAGTAACGGCCGCCAGTGTGCTCTAAGGCCACC  
TGCAGGGCTACTGCTGCTCCGGCCACTGCCTGAGACTCACCTTGCTGGAACGTG  
AGCCTCGGCTTCTGTCATCTCTGTGGCCTCTGTCGCTTCTGTCGCTGTCCCCCTTN  
AGTCCCTGAGCCCCGCGAGCCCCGGGCCGCACACGCGAATGGGCGGCACGGCC  
AGGGCGCTGGGTGGCGGTGGGGCTGGGCGTCGTGGGGCTGCTGTGCGCTGTGCT  
CGGTGTGGTTATGATCCTCGTGATGCCCTCGCTCATCAAACAGCAGGTACTGA  
AGAATGTCCGCATAGACCCAGCAGCCTGTCTTTGCAATGTGGAAGGAGATC  
CCTGTACCCTTCTACTTGTCCGTCTACTTCTTCGAGGTGGTCAATCCCAGCGAG  
ATCCTAAAGGGTGAGAA

BLASTX 1.3.9MP

[Build

Reference: Gish, Warren and David J. States (1993). Identification of protein coding regions by database similarity search. Nature Genetics 3:266-72.  
Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman (1990). Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Notice: statistical significance is estimated under the assumption that the equivalent of one reading frame in the query sequence codes for protein and that significant alignments will involve only coding reading frames.

Query= TITLE phasr3.seq  
(447 letters)


Translating both strands of query sequence in all 6 reading frames

Database: Non-redundant PDB+SwissProt+PIR+SPupdate+GenPept+GPupdate,  
EDT  
96,634 sequences; 27,090,059 total letters.

Searching.....done

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Poisson Probability P(N)	N	
sp P27615 LIM2_RAT	LYSOSOME MEMBRANE PROTEIN II (L...	+2	114	1.1e-08	1
pir JQ1523 JQ1523	lysosomal membrane 85K sialogly...	+2	109	6.3e-08	1
sp P10284 HM26_MOUSE	HOMEBOX PROTEIN HOX-2.6. >pir ...	-2	61	2.4e-06	2
sp P16671 CD36_HUMAN	PLATELET GLYCOPROTEIN IV (GPIV)...	+2	94	1.1e-05	1
gp L06850 HUMCD36B_1	antigen CD36 [Homo sapiens]	+2	94	1.1e-05	1
gp L19658 RATFAT_1	FAT gene product [Rattus norveg...	+2	92	2.3e-05	1
pir A43932 A43932	mucin - human (fragment)   0.0 ...	-1	60	3.8e-05	2
pir B60492 B60492	homeotic protein Hox B4 - human...	-2	57	4.0e-05	2
sp Q01200 PRIA_LENED	PRIA PROTEIN. >pir S23106 S2310...	-1	62	5.6e-05	2
pir S12968 S12968	Acrosin, sperm - Pig #EC-number...	-2	59	6.7e-05	2
gp L23108 MUSCDANTI_1	CD36 antigen [Mus musculus]	+2	88	9.0e-05	1
pir A45106 A45106	mucin - human (fragment)   0.0 ...	-1	60	9.2e-05	2
pir S31976 S31976	Cvx peptide - Rat   0.0 0.0 0.0...	-3	57	0.00012	2
gp Z16406 MOX2A_1	Mox-2 [Mus musculus]	-3	57	0.00012	2
gp Z17223 RNGAXMR_1	Gax peptide [Rattus norvegicus]	-3	57	0.00012	2
sp P13983 EXTN_TOBAC	EXTENSIN PRECURSOR (CELL WALL H...	-2	56	0.00024	2
pir G60110 G60110	repetitive protein antigen 69/7...	-2	81	0.00035	1
gp M14721 MUSFGNAA_1	Mouse epidermal profilaggrin mR...	-3	71	0.0044	1
pir B36664 B36664	S59/4 homeotic protein - fruit ...	-3	76	0.0080	1

>sp|P27615|LIM2\_RAT LYSOSOME MEMBRANE PROTEIN II (LIMP II) (85 KD LYSOSOMAL  
MEMBRANE SIALOGLYCOPROTEIN) (LGP85). >pir|A41180|A41180 74k  
lysosomal membrane protein LIMP - rat | 0.0 0.0 0.0 0.0 0.0  
>pir|JH0241|JH0241 85K lysosomal membrane sialoglycoprotein - rat |  
0.0 0.0 0.0 0.0 0.0 >gp|D10587|RATLGP85\_1 LGP85 [Rattus sp.]  
>gp|M68965|RATLIMP2\_1 lysosomal membrane protein [Rattus  
norvegicus]  
Length = 478



```
>pir|JQ1523|JQ1523 lysosomal membrane 85K sialoglycoprotein precursor - human |
0.0 0.0 0.0 0.0 0.0 0.0 >gp|D12676|HUMHLGP85_1 85kDa human lysosomal
sialoglycoprotein [Homo sapiens]
Length = 478
```

```
>sp|P10284|HM26_MOUSE_HOMEOBOX_PROTEIN_HOX-2.6. >pir|A31757|A31757 homeotic
protein Hox 2.6 - mouse | 0.0 0.0 0.0 0.0 0.0 >gp|M36654|MUSHOX26_1
Mouse homeo box 2.6 (Hox-2.6) mRNA, complete cds. [Mus musculus]
Length = 250
```

```
>sp|P16671|CD36_HUMAN PLATELET GLYCOPROTEIN IV (GPIV) (GPIIIB) (CD36 ANTIGEN).
>pir|A30989|A30989 CD36 protein - human | 0.0 0.0 0.0 0.0 0.0
>qp|M24795|HUMAN TCD36 1 Human CD36 antigen mRNA, complete cds.
```

[Homo sapiens] >gp|M98398|HUMCD3613\_1 antigen CD36 [Homo sapiens]  
>gp|M98399|HUMCD3621\_1 antigen CD36 [Homo sapiens]  
Length = 472

Plus Strand HSPs:

Score = 94 (45.5 bits), Expect = 1.1e-05, P = 1.1e-05  
Identities = 18/64 (28%), Positives = 36/64 (56%), Frame = +2

Query: 245 VVGLLCAVLGVVMILVMPSLIKQQLKNVRIDPSSLSFAMWKEIPVPFYLSVYFFEVDVNP 424  
V+G + AV G +++ V LI++ + K V ++ +++F W + Y + F+V NP  
Sbjct: 14 VIGAVLAVFGGILMPVGDLLIQTIKKQVVLEEGTIAFKNWVKGTGEVYRQFWIFDVQNP 73

Query: 425 SEIL 436

E++

Sbjct: 74 QEVM 77

>gp|L06850|HUMCD36B\_1 antigen CD36 [Homo sapiens]  
Length = 472

Plus Strand HSPs:

Score = 94 (45.5 bits), Expect = 1.1e-05, P = 1.1e-05  
Identities = 18/64 (28%), Positives = 36/64 (56%), Frame = +2

Query: 245 VVGLLCAVLGVVMILVMPSLIKQQLKNVRIDPSSLSFAMWKEIPVPFYLSVYFFEVDVNP 424  
V+G + AV G +++ V LI++ + K V ++ +++F W + Y + F+V NP  
Sbjct: 14 VIGAVLAVFGGILMPVGDLLIQTIKKQVVLEEGTIAFKNWVKGTGEVYRQFWIFDVQNP 73

Query: 425 SEIL 436

E++

Sbjct: 74 QEVM 77

>gp|L19658|RATFAT\_1 FAT gene product [Rattus norvegicus]  
Length = 472

Plus Strand HSPs:

Score = 92 (44.5 bits), Expect = 2.3e-05, P = 2.3e-05  
Identities = 18/65 (27%), Positives = 36/65 (55%), Frame = +2

Query: 245 VVGLLCAVLGVVMILVMPSLIKQQLKNVRIDPSSLSFAMWKEIPVPFYLSVYFFEVDVNP 424  
V+G + AV G +++ V LI++ + + V ++ +++F W + Y + F+V NP  
Sbjct: 14 VIGAVLAVFGGILMPVGDLLIEKTIKREVLEEGTIAFKNWVKGTGTVYRQFWVFDVQNP 73

Query: 425 SEILK 439

E+ K

Sbjct: 74 EEVAK 78

>pir|A43932|A43932 mucin - human (fragment) | 0.0 0.0 0.0 0.0 0.0  
>gp|M74027|HUMMUC2A\_1 mucin [Homo sapiens]  
Length = 573

Minus Strand HSPs:

Score = 60 (29.1 bits), Expect = 1.4, P = 0.74  
Identities = 12/21 (57%), Positives = 14/21 (66%), Frame = -1

Query: 279 TTPSTAHSPTTPSPTATQRP 217  
TPS ++ TTPSPT T P  
Sbjct: 377 TTPSPPTTMTTPSPTTTPSP 397

Score = 58 (28.1 bits), Expect = 3.8e-05, Poisson P(2) = 3.8e-05  
Identities = 12/20 (60%), Positives = 14/20 (70%), Frame = -1

Query: 285 IITTPSTAHSPTTPSPTAT 226  
I TTPS ++ TTPSPT T  
Sbjct: 343 ITTTPSPPTTMTTPSPTTT 362

>pir|B60492|B60492 homeotic protein Hox B4 - human | 0.0 0.0 0.0 0.0 0.0  
Length = 251

Minus Strand HSPs:

Score = 57 (27.8 bits), Expect = 2.9, P = 0.95  
Identities = 12/21 (57%), Positives = 12/21 (57%), Frame = -2

Query: 266 QRTAAPRRPAPPPPSALAVPP 204  
QR AA R P PPPP PP  
Sbjct: 63 QRYAACRDPGPPPPPPPPPP 83

Score = 56 (27.3 bits), Expect = 4.0e-05, Poisson P(2) = 4.0e-05  
Identities = 11/20 (55%), Positives = 12/20 (60%), Frame = -2

Query: 254 APRRPAPPPPSALAVPPMSR 195  
+PR PAPPP AL P R  
Sbjct: 90 SPRAPAPPPAGALLPEPGQR 109

>sp|Q01200|PRIA\_LENED PRIA PROTEIN. >pir|S23106|S23106 priA protein - Shiitake  
mushroom | 0.0 0.0 0.0 0.0 0.0 0.0 >gp|X60956|LEPRIA\_1 priA gene  
product [Lentinus edodes]  
Length = 258

Minus Strand HSPs:

Score = 62 (30.0 bits), Expect = 0.61, P = 0.46  
Identities = 13/31 (41%), Positives = 18/31 (58%), Frame = -1

Query: 318 TCCLMSEGITRIITTPSTAHSPTTPSPTAT 226  
TCCL + TPS+AH + T SP++T  
Sbjct: 90 TCCLPKWPTSTPTPTPSSAHTSTHTSPSST 120

Score = 56 (27.1 bits), Expect = 5.6e-05, Poisson P(2) = 5.6e-05  
Identities = 13/33 (39%), Positives = 16/33 (48%), Frame = -1

Query: 276 TPSTAHSPTTPSPTATQRPGRAAHVACAARAR 178  
TPS+ +TP P+AT G H A AR  
Sbjct: 143 TPSSPSKPSSTPKPSATPNKGNHGHYKRAHVAR 175

>pir|S12968|S12968 Acrozin, sperm - Pig #EC-number 3.4.21.10 | 0.0 0.0 0.0 0.0

0.0  
Length = 374

Minus Strand HSPs:

Score = 59 (28.8 bits), Expect = 1.6, P = 0.79  
Identities = 14/48 (29%), Positives = 24/48 (50%), Frame = -2

Query: 251 PRRPAPPPPSALAVPPMSRVRPGLAGLRDSRGTATEATEATEMTEAEA 108  
P++ + PP AL+ + ++ L G S G + TE T++ E A  
Sbjct: 326 PQQVSAKPPQALSFAKLQQLIEALKGTAFSSGRSYETETTDLQELPA 373

Score = 56 (27.3 bits), Expect = 6.7e-05, Poisson P(2) = 6.7e-05  
Identities = 11/21 (52%), Positives = 12/21 (57%), Frame = -2

Query: 266 QRTAAPRRPAPPPPSALAVPP 204  
Q + PR PAPPPP PP  
Sbjct: 294 QPGSRPRPPAPPPPPPPPPPP 314

>gp|L23108|MUSCDANTI\_1 CD36 antigen [Mus musculus]  
Length = 473

Plus Strand HSPs:

Score = 88 (42.6 bits), Expect = 9.0e-05, P = 9.0e-05  
Identities = 17/65 (26%), Positives = 35/65 (53%), Frame = +2

Query: 245 VVGLLCAVLGVVMILVMPSLIKQQVLKNVRIDPSSLSFAMWKEIPVPFYLSVYFFEVVNP 424  
V+G + AV G +++ V LI++ + + V ++ + +F W + Y + F+V NP  
Sbjct: 15 VIGAVLAVFGGILMPVGDMLIEKTIKREVVLEEGTTAFKNWVKGTGTVYRQFWIFDVQNP 74

Query: 425 SEILK 439  
++ K  
Sbjct: 75 DDVAK 79